

Bioinformatics Core Unit

64

Scientific Report 2010  CNIO



David G. Pisano

Unit Head

Born in 1970 in Oviedo, Spain, Pisano graduated in Biology from the *Universidad de Oviedo*. His postgraduate work involved the application of numeric techniques to analyse stability data and the development of mathematic language interpreters to process HPLC signals.

In 1998 he was recruited by Thermo Fisher Scientific and was shortly after appointed as Services Delivery Manager for Spain, Portugal and Latin America. At Thermo Pisano gained international experience in providing professional IT services for research, analytical and industrial sectors.

In 2001 he joined the molecular diagnostics company *Genómica* (Zeltia), Madrid, as Head Bioinformatician where he worked at the Functional Genomics Unit focusing on development of cancer biomarker discovery tools using DNA array technology.

Pisano moved to the *Centro Nacional de Biotecnología* (CNB-CSIC), Madrid, in 2004 as Technical Manager for the Spanish National Bioinformatics Institute (INB). He was in charge of establishing a nation-wide Bioinformatics platform to support large-scale genome projects.

He joined the CNIO as Head of the Bioinformatics Unit in 2006.

Summary

The biomedical research landscape has witnessed significant changes over the last decades. The amount of molecular data produced even in the smallest laboratory is exponentially increasing and the Molecular Biology field is starting to emerge as a quantitative science. Access to computational and statistical tools to analyse the data in a timely and effective way and obtain biologically sound interpretations is critical for scientists.

In this context, the Bioinformatics Core Unit is dedicated to helping CNIO researchers understand the biological processes underlying cancer by applying state-of-the-art computational and numerical methodologies to the study of large-scale experimental datasets.

Main Objectives

- Provide custom bioinformatics support to CNIO researchers through bioinformatic analysis services and scientific computing technologies
- Develop innovative hardware and software-based computational solutions to address problems related to the storage, management, and analysis of high throughput biological data
- Improve the interpretation of complex numerical data through the provision of custom advice and training in novel Bioinformatics methods





Technicians: Eduardo Andrés, Ángel Carro, Gonzalo Gómez and Osvaldo Graña.



Figure: ChIP-seq allows *in vivo* determination of the locations of DNA-binding proteins at genomic level. The figure shows the extra-telomeric distribution of 30,398 binding sites for shelterin RAP1 in the mouse genome. Each bar shows the number of detected sites in a 1 Mbp window (grey) including those that are in the neighbourhood of a known gene (orange).

Highlights

This year the focus of the Unit has significantly shifted to provide support for Next Generation Sequencing (NGS) experiments. This technology allows deep sequencing of DNA or RNA molecules and requires substantial bioinformatics resources to manage, analyse and interpret the results.

During 2010 the Unit set up a number of NGS data analysis pipelines for RNA-seq, ChIP-seq (Figure), exome resequencing for the discovery of somatic and germline variations, and MethylC-seq. Activities included evaluation and benchmarking of multiple NGS algorithms, operation and implementation of software tools to interpret NGS experiments, and the development of custom software pipelines.

The Bioinformatics Core Unit also operates the CNIO's scientific computing facilities. In 2010 we upgraded several computational clusters to meet the increasing demands for all NGS methods. The current computing farm hosts more than 1200 CPUs, has 2700GB of available RAM memory and can store up to 35TB of data. We also began exploring remote computational and storage solutions (cloud computing) to reduce power, cooling and costs associated with hardware ownership.

We have also continued to provide Bioinformatics assessment and training to CNIO researchers on data analysis oriented toward array-based and sequence analysis experiments. As requested by researchers, we developed custom software solutions for miRNA target determination, designed genomic variation validation assays in batch, and carried out rich data visualisation and quality control at genome-scale.

Publications

Martinez P, Thanasoula M, Carlos Ana R, Gómez-Lopez G, Tejera AM, Schoeftner S, Dominguez O, Pisano DG, Tarsounas M, Blasco MA (2010). Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat Cell Biol* 12, 768-780.

Di Lisio L, Gómez-López G, Sánchez-Beato M, Gómez-Abad C, Rodríguez ME, Villuendas R, Ferreira BI, Carro A, Rico D, Mollejo M, Martínez MA, Menárguez J, Díaz-Alderete A, Gil J, Cigudosa JC, Pisano DG, Piris MA, Martínez N (2010). Mantle cell lymphoma: transcriptional regulation by microRNAs. *Leukemia* 24, 1335-1342.

Carro A, Rico D, Rueda OM, Díaz-Urriarte R, Pisano DG (2010). waviCGH: a web application for the analysis and visualization of genomic copy number alterations. *Nucleic Acids Res* 38, W182-W187.

Pettifer S, et al. (2010). The EMBRACE web service collection. *Nucleic Acids Res* 38, W683-W688.

Carilla-Latorre S, Gallardo ME, Annesley SJ, Calvo-Garrido J, Graña O, Accari SL, Smith PK, Valencia A, Garesse R, Fisher PR, Escalante R (2010). MidA is a putative methyltransferase that is required for mitochondrial complex I function. *J Cell Sci* 123, 1674-1683.

López-Jiménez E, et al. (2010). Research Resource: transcriptional Profiling Reveals Different Pseudohypoxic Signatures in SDHB and VHL-Related Pheochromocytomas. *Mol Endocrinol* 24, 2382-2391.

Orenes E, Barderas R, Rico D, Casal I, González-Pisano D, Navajo J, Algaba F, Piulats JM, Sanchez-Carbayo M (2010). Serum and tissue profiling in bladder cancer combining protein and tissue arrays. *J Proteome Res* 9, 164-173.

García-Jiménez B, et al. (2010). Inference of functional relations in predicted protein networks with a machine learning approach. *PLoS One* 5, e9969.

Ferreiro I, Joaquin M, Islam A, Gomez-Lopez G, Barragan M, Lombardía L, Domínguez O, Pisano DG, Lopez-Bigas N, Nebreda AR, Posas F (2010). Whole genome analysis of p38 SAPK-mediated gene expression upon stress. *BMC Genomics* 11, 144.

Zajac M, Gomez G, Benitez J, Martínez-Delgado B (2010). Molecular signature of response and potential pathways related to resistance to the HSP90 inhibitor, 17AAG, in breast cancer. *BMC Genomics* 3, 44.

Glez-Peña D, Reboiro-Jato M, Domínguez R, Gómez-López G, Pisano DG, Fdez-Riverola F (2010). PathJam: a new service for integrating biological pathway information. *J Integr Bioinform* 7, 147.